Patent Office
Research and Development
Reports . . . . . No. 19

REVISED STEROID SEARCH SYSTEM CODING MANUAL

Prepared by

Julius Frome

Patent Research Specialist

Office of Research and Development

Patent Office

(Reprinted July 1962)

U. S. DEPARTMENT OF COMMERCE
Frederick H. Mueller
Secretary

PATENT OFFICE
Robert C. Watson
Commissioner

For sale by the U. S. Department of Commerce, Washington 25, D. C. Price 25 cents.

CONTENTS

ACKNOWLEDGMENTS

Appreciation is expressed for the aid and cooperation of the full steroid committee, the Ad Hoc Committee consisting of Dr. Howard T. Bonnett, Dr. Helen Ginsberg, Miss Winifred Sewell and Mr. T. E. Arther; examiners of Mechanized Examining Division A - Walter A. Modance, Lewis Gotts, Elbert L. Roberts, Thomas J. Morgan; and Martin A. Gannon and Paul T. O'Day of Research and Development. These people have given unstintingly of their time and technical knowledge.

REVISED STEROID SEARCH SYSTEM CODING MANUAL

This manual supersedes Patent Office Research and Development Report No. 7, *A Punched Card System for Searching Steroid Compounds* and Report No. 11, *A Manual for Coding Steroids*. The present manual reflects the results of two years of experience by the Patent Office and the Industry in the use of a punched card system for conducting searches in the steroid art.

SCOPE OF THE ART IN THE SYSTEM

The system is limited to the steroid art. The patents included in the system are those which are classified in Class 260, sub-classes 239.5, 239.55, 239.57, 397, 397.1, 397.2, 397.35, 397.3, 397.4, 397.45, 397.47, and 397.5, and includes some steroid containing patents gleaned from other sub-classes. Only those steroids disclosed in the patents which meet the definitions of these class and sub-classes of the Patent Office classification are included in the system.

Every compound which is coded must contain the steroid nucleus shown in Figure 1 in which the positional locations of nuclear substitution are identified by numbers assigned to each position. The seco and homo steroids

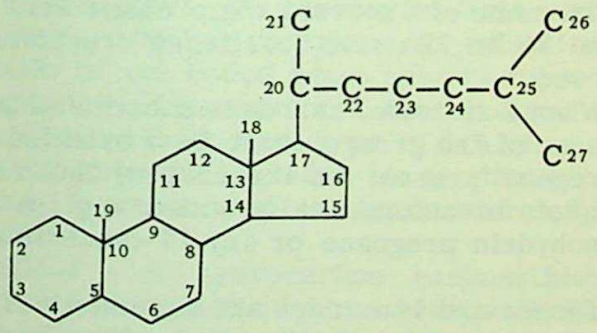


Figure 1

are excluded. Also excluded are steroids classified elsewhere according to the Rule of Superiority in classification.

Positional locations 18 through 23 are variables in the sense that they may not be present in a particular compound (e.g., androstanes). The locations exist only when they are occupied by carbon atoms.

RELATIONSHIP OF CODES AND PUNCH-CARD FORMAT

Every significant term or substituent of the chemical steroid formula is called a Descriptor.

A descriptor is defined by a single set of numbers designating a particular column and row of the standard 80-column IBM punched card. Conversely, each column-row location on the IBM is a fixed allocated space for a particular descriptor. The card is divided into two fields, columns 1 through 50 constituting the field for 2A terms and columns 51-65 constituting the 2B terms. The more specific columns and rows for each descriptor can be seen from the steroid code sheet which contains the column and row number for each descriptor.

The definition of each of the descriptors on the steroid code sheet follow infra.

REMARKS ABOUT CODING

Broadly speaking, the arrangement of descriptors on the card is according to three general types of information, namely:

(1) The "2-A" field (columns 1-50) in which 24 structural concepts plus miscellaneous may be recorded as being present in any of positions 1-23, and "Ex." (The "Ex" designation is an abbreviation for "Exo" and is used to code substituents which are not directly connected to the 1-23+ carbon atoms of the steroid nucleus. See Appendix A.)

(2) The "2-B" field (columns 51-65) which is used primarily for recording structural concepts but without defining their position on the nucleus.

For purposes of coding patents all of the steroid formulae disclosed within a patent are combined to produce a single synthetic formula (this process is called "compositing"). Each functional group appearing in this synthetic formula together with its position is punched into the 2-A and 2-B fields of the code format.

(3) Document identification and a card serial number.

Columns 66-69 are used to record a number which is a chronologically assigned serial number to the cards in the patent deck. This number will parallel the patent numbers, except

for reissue patents. Subscribers to the patent deck can use this number to check their decks for completeness.

Column 70 is used to record the patent office classification of patents. See Appendix C for the code used and the type of steroids in each category. The coding in this column may provide a useful search tool, but it should be borne in mind that the classification assigned by the Patent Office is based upon *claimed* subject matter, rather than subject matter *disclosed*.

The Column 71-80 field is used to record the literature reference or patent number. See Appendix D for the coding used in these columns.

It is important to note that where chemistry and information retrieval are at variance, for the purpose of this manual, information retrieval takes precedence.

INSTRUCTIONS FOR CODING

1. Read the document for comprehension of the subject matter.

2. Encircle all of the codes representing both the 2A and 2B terms found in the patent on the coding data sheet in accordance with the principles of multiple coding and composite coding.

3. Extract all pertinent terms disclosed in the patent. The title, text, and claims of the patent are all parts of the disclosure for extraction and coding purposes. Chemical configurations disclosed as possible substituents as well as those more specifically disclosed are extracted and coded.

4. Have the coded information verified by a second individual. Note that the coded data sheet represents a composite of all the substituents disclosed for the steroid nucleus in the given patent.

DEFINITION OF CODE TERMS

1. 2-A Terms

= This symbol designates a carbon-to-carbon double bond involving positions 1-23. In coding the positional location of a double bond, the lower number of the pair of position numbers is recorded. Thus for $\Delta^{4,5}$, position 4 is coded, and for 17(20), code 17. The 2-A field punches will not be definitive for positions 1, 5, 8, 9, 13 and 20; distinction between the two variants in each of these positions is provided in the 2-B field. (Col. 51) The unsaturated linkage of methylene groups connected to the steroid nucleus is coded as "exo" double bond only. Position 10 in the 2-A field is not used.

An aromatic A ring is coded 1, 3, 5 in the 2-A field, and in addition is coded 1(2), 5(10) and aromatic A in the 2-B field.

Ring saturation is also indicated in the 2-B terms. Double bonds that are common to two rings are treated as follows:

5(10) - A ring
8(9) - C ring
13(14) - C ring

2. -H

This symbol is applied only when the hydrogen atom is attached to the 10, 13, or 20 carbon atom to indicate the absence of a methyl group in the 18, 19, or 21 position. For example, when there is a 17 formyl group, -H is coded at the 20 position. It is not used for a 17-methyl group.

NOTE: The hydrogen actually has to be present to use this descriptor. Thus, in the case of steroids containing an aromatic A ring which has neither a methyl nor a hydrogen attached to position 10, the 10-H code is not used.

3. α or allo

This term represents a particular orientation at an asymmetric steroidal carbon.

The α descriptor is implied and coded wherever the configuration is known from the trivial name of a steroid, e.g., "cholic acid" is coded 3α 7α 12α and "cortisone" receives a 17α code.

When a steroidal carbon is substituted such that one of the groups must be α oriented (no hydrogens present on the carbon) the α descriptor is automatically coded, e.g., a 20-cyanohydrin pregnane or any 17-disubstituted steroid.

The 9α and 14α codes are assumed and not coded unless the substituent is other than H.

Note that i compounds have a specific 2-B punch and α or allo is not used to identify i compounds. The i compounds are located in 2-B under "General."

However, a word of caution in using the α descriptor. Since, in the absence of a specific statement in the document, the coder may not have known whether a particular substituent was α or not, the descriptor may have been omitted in some cases. It should not be used when a complete search is desired.

4. β (beta)

This code is only used when the document *specifically states* the β configuration. It is

never implied and therefore may not be relied on for complete searches.

The 5-B code is not used unless the substituent is other than hydrogen.

NOTE: Since coding the α and β descriptors is sometimes arbitrary, they should not be used if a complete search is desired.

5. Miscellaneous

This designation is used for any group not provided for by any other 2-A term.

Examples of groups in this category are -O-Na, azides (consider $-\overset{\overset{O}{\|}}{C}-N_3$ as a unit), and Grignard intermediates.

Rule of Superiority. The Misc. code is used only if no other designation is applicable. (Note: In cases of doubt between C-sub and Misc., the C-sub code takes precedence.)

6. -OH

This represents the hydroxy group.

7. =O

This represents the keto group.

8. -O-Acyl

This designation refers to an ester group attached to the steroid through the -O- atom of the group. It is further defined in the 2-B terms as follows:

Carboxylic - A carboxylic acid radical.

Poly - A polycarboxylic acid radical (exo-COOR is not coded when poly is recorded).

Unsat - An unsaturated acid radical. The double bond is not coded exo in 2-A. Aromatic unsaturation is excluded (e.g., the benzoyl group is not coded as an unsaturated acid radical).

Subst - A hydrocarbon carboxylic acid radical having a non hydrocarbon substituent. This term is applicable to poly, unsaturated, aromatic, and aliphatic radicals. The carboxyl group of a polycarboxylic acyl is excluded from this category. Examples are:

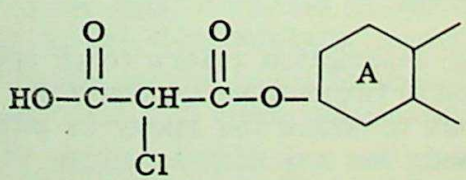


Figure 2

coded as carboxylic, poly, subst, halo cont., "exo" halo, and "Cl" in 2 B. (53-0, 1, 3; 54-8; 25-Ex; 56-8)

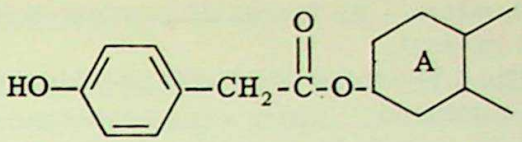


Figure 3

coded as carboxylic, aromatic, subst, O-cont., and "exo" OH. (53-0, 3, 4; 54-7; 11-Ex)

Aromatic - A carboxylic acid radical containing an aromatic hydrocarbon ring.

Aliphatic - An aliphatic, non-aromatic carboxylic acid radical.

St. Chain - A straight chain aliphatic acid radical with no non-hydrocarbon substituents (formate and acetate included).

Cycloalkyl - An aliphatic hydrocarbon carboxylic acid containing a cycloalkyl group (e.g., cyclopentylpropionate).

Branched - An aliphatic branched hydrocarbon carboxylic acid radical.

Heterocyclic - A heterocyclic containing carboxylic or inorganic acid radical (e.g., nicotinate).

Inorganic Acyl - Includes the acyl group of any inorganic acid with the exception of the halogen acids. Note: This term includes the acyl radicals derived from carbonic acid, its alkyl esters, and its inorganic derivatives (e.g., phosgene), but does *not* include acyl radicals derived from carbamic or xanthic acids - these are coded as substituted aliphatic carboxylic acid radicals.

Phosphorus A - An inorganic phosphorous - containing acyl radical.

SO_4 (all S) - This term includes all inorganic sulfur-containing acyl radicals. Examples are the mesyl and tosyl radicals.

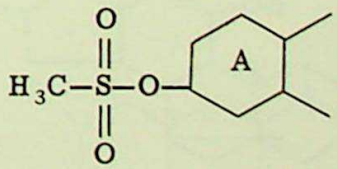


Figure 4

coded inorg, SO_4, aliphatic, S-cont. and O-cont. (53-11; 54-0, 3, 6, 7)

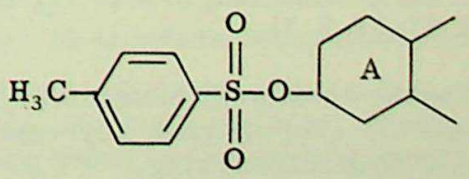


Figure 5

coded inorg, SO_4, aromatic, S-cont., and O-cont. (53-11, 54-0, 4, 6, 7)

Osmium - An inorganic osmium-containing acyl radical.

The 4(5) osmate of progesterone exemplifies this term:

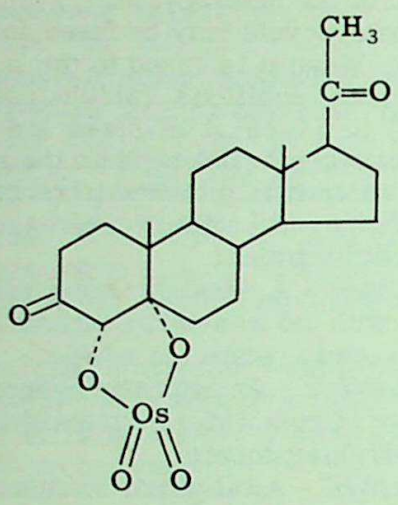


Figure 6

The osmate radical is coded as 4, 5-0-acyl in 2-A and as inorganic, osmium, and O-containing in 2-B. (15-4, 5; 53-11; 54-1, 7)

Boron - An inorganic boron-containing acyl radical.

The 16, 17-cycloborate of 16α hydroxyhydrocortisone exemplifies this term.

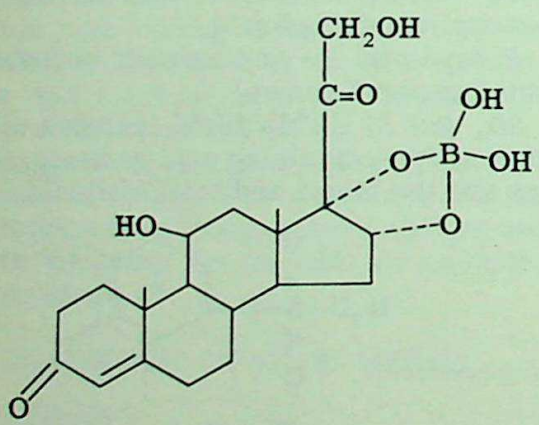


Figure 7

The cycloborate radical is coded as exo-OH and 16, 17 O-acyl in 2-A and as inorganic, boron, and O-containing in 2-B. (11-Ex; 16-16, 17; 53-11; 54-2, 7)

Aliphatic - An acyclic inorganic acyl radical.

Aromatic - An inorganic acyl containing an aromatic hydrocarbon ring.

N, S, O, and Halo-containing are coded when an -O-Acyl radical, either organic or inorganic, contains any one of these elements, excluding the oxygen contained in the carboxyl groups of organic acyl radicals. The O-containing descriptor is coded in the case of SO_4, PO_4 and osmates. It is not coded for polycarboxylic acids.

Miscellaneous - Any O-Acyl not provided for above.

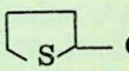
Note: As a result of the multiple coding principle, all applicable terms are applied to a particular compound.

9. -O-R

This defines the ether linkage, i.e., R-O-R, where one R is a carbon which is a part of the steroid nucleus (except for "Exo") and the other R is a substituent (aliphatic, aromatic, cycloaliphatic, or heterocyclic) which is attached to the oxygen atom independently through a carbon atom and which is further defined in 2-B by the following terms:

R = hydrocarbon, e.g., CH_3O-A

R = N group, e.g., NH_2-CH_2-O-A

R = S group, e.g.,  CH_2-O-A

R = O group, e.g., HO-CH_2-O-A

R = halo group, e.g., F_3-C-O-A

R = other

Note: Because they are specifically provided for elsewhere acetals and ketals are not coded under O-R.

10. Epoxy

This description refers to an epoxy group attached to two nuclear carbon atoms. The two positions to which the epoxy is attached are recorded.

Since this term is closely related to the O-hetero category the following rules are applied:

I All carbons linked by a single oxygen atom are coded as epoxy unless the oxygen is a member of a lactone group in which case "epoxy" is not used. E.g.:

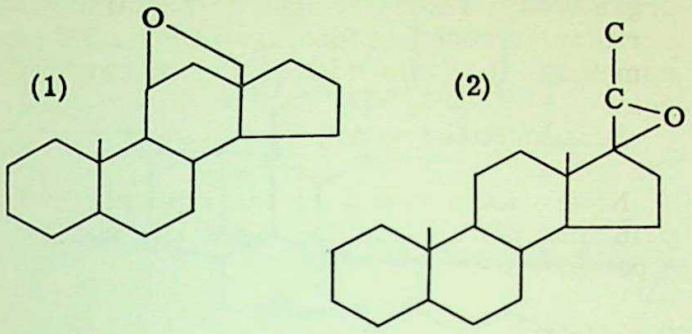


Figure 8

are coded (1) 11, 18 epoxy and (2) 17, 20 epoxy.

II Epoxides of *non-adjacent* carbons involving one or two of the 20, 21, 22 and/or 23 + carbons are also coded as O-hetero. See the O-hetero section for the proper coding procedure in such cases. E.g.:

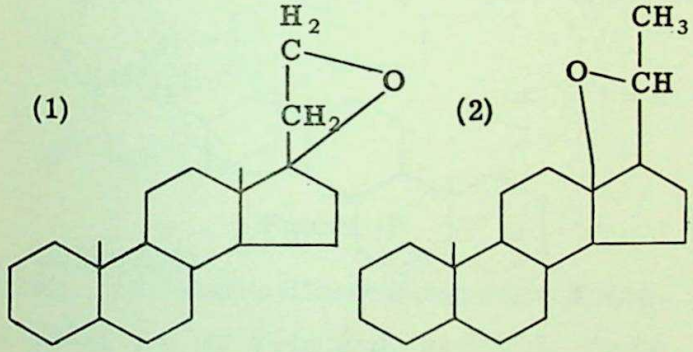


Figure 9

are coded (1) 17, 21 epoxy; 17 - O-hetero, spiro, misc and (2) 18, 20 epoxy; 13, 17 - O-hetero-fused-furan.

11. Ketal:

This designation refers to the reaction product of a keto or aldehyde group with an alcohol to give either a cyclic or non-cyclic ketal or acetal.

The term ketal includes thioketal, semi-thioketal, acetal, thioacetal and hemithioacetal.

Non cyclic ketals (acetals) should also be coded under the 2-B term "Bis-substituents."

Cyclic oxygen ketals are not coded in the O-hetero category in 2-A or in 2-B. Cyclic thioketals, including hemithioketals are not coded under S-hetero in 2-A but are coded as thioketal in the 2-B S-hetero column.

Non cyclic ketals (acetals) are not coded under O-R or S-R in 2-A or 2-B.

12. O-Hetero

This designation refers to an oxygen-containing heterocyclic group. The heterocycle can be attached to the steroid nucleus through any atom of the heterocycle.

The heterocycle may be fused, independent, or spiro. When it is fused to the nucleus, the two nuclear positions (which need not be adjacent) to which it is fused are recorded. When it is spiro or independent the one position through which it is attached is recorded.

Spiro

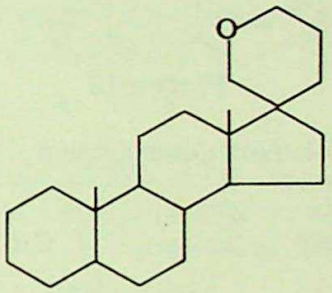


Figure 10

Code: 17-O-Hetero; pyranyl; spiro

Fused

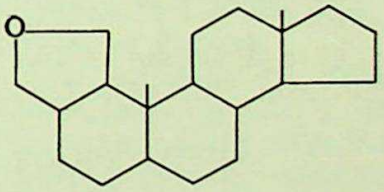


Figure 11

Code: 1,2-O-hetero; furan; fused

Independent

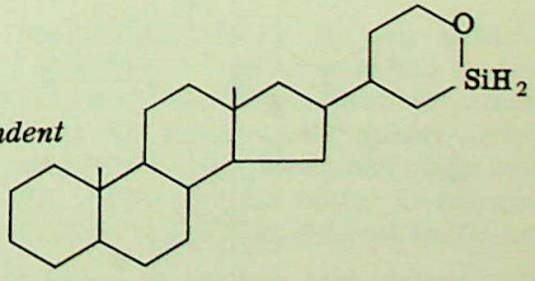


Figure 12

Code: 16-O-hetero; independent; misc.

Acid and anhydride adducts (e.g., maleic anhydride adducts), epoxides, cyclic oxygen containing ketals and steroidal sapogenins are specifically excluded from the 2-A term O-hetero.

When the 18-23+ side chain carbons are members of the hetero ring, the point of

attachment to the steroid ring is recorded unless the hetero ring is independently connected to a side chain carbon. E.g.:

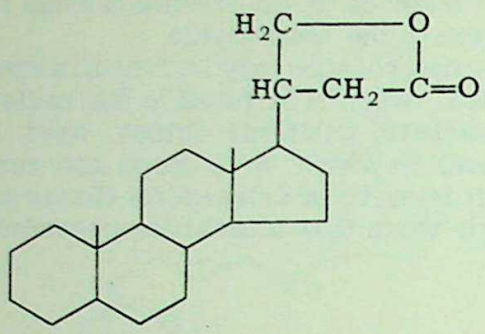


Figure 13

Code: 17-O-hetero; independent; lactone; furan

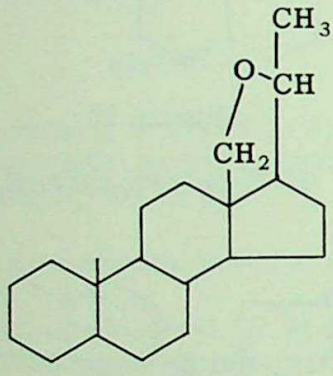


Figure 14

Code: 13, 17-O-hetero; fused; furan
(Note: Also code 18, 20-epoxy)

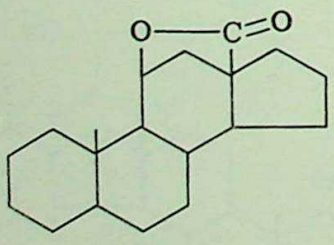


Figure 15

Code: 11, 13-O-hetero; fused; lactone; furan

The O-hetero is further defined in 2-A as follows:

Morpholine - this is also coded as N-hetero

Furan - includes saturated and unsaturated forms and also 5-member lactones

Lactone - E.g.:

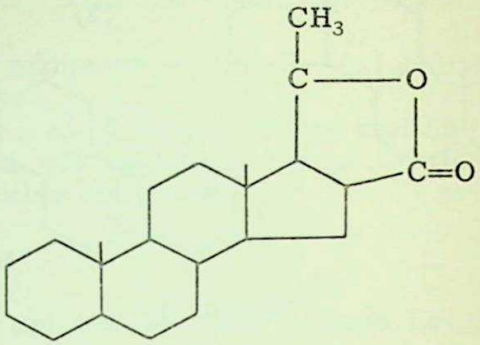
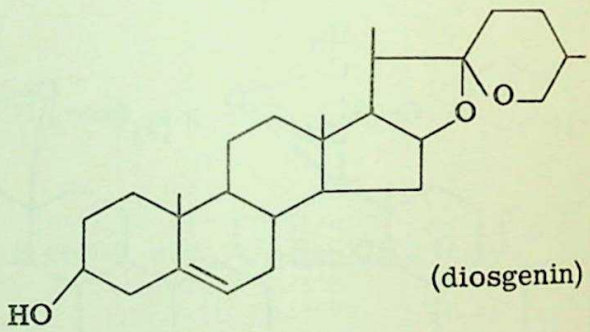


Figure 16

is coded 16, 17 O-hetero, furan, lactone, fused.

Spirostane - includes both normal and iso. E.g.:



(diosgenin)

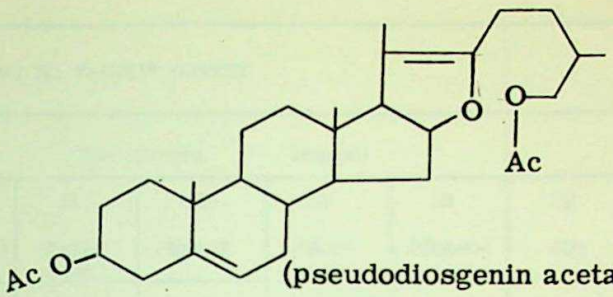
Figure 17

The above formula is coded as "spirostane" - (no other O-hetero descriptors in 2-A or 2-B are used)

Sub in O-Spiro Ring - spirostanes substituted in the oxygen rings by a non-hydrocarbon substituent. The substituent is also coded as "exo."

NOTE: The 2-A designations 16, 17-O-hetero and 2-B fused are not recorded for sapogenins or for pseudo-sapogenins.

Pseudosapo - the pseudosapogenins are derived from steroidal sapogenins by treatment with acid anhydrides. The free hydroxy analog is included in this descriptor. The side chain hydroxy or acetate is not coded "exo." The double bond in the fused O-hetero ring is not coded as "exo" - it is included in the term "pseudosapogenin." E.g.:



(pseudodiosgenin acetate)

Figure 18

is coded as "pseudosapogenin."

Acetonide - cyclic acetal or ketal which is the reaction product of an aldehyde or ketone with 2 hydroxy groups attached to the steroid nucleus. It is also coded as "fused." E.g.:

The acetonide portion of

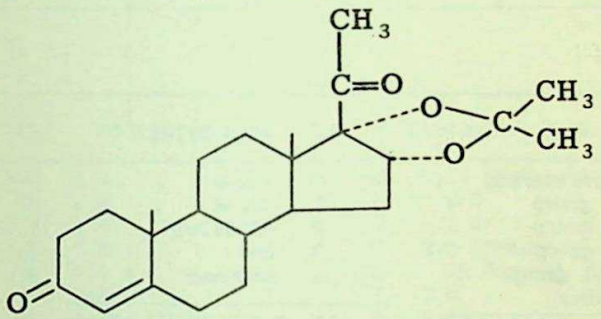


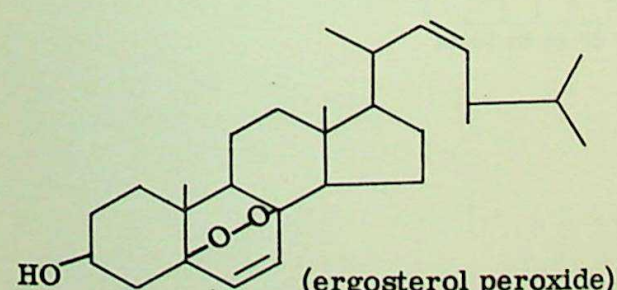
Figure 19

(16α, 17α isopropylidenedioxyprogesterone)

is coded 16, 17 O-hetero, acetonide, fused.

NOTE: The methyl groups may be replaced by hydrogen or any other substituent (hydrocarbon or substituted hydrocarbon, heterocyclic, halogen, etc.) or by substituents which themselves form a ring.

Peroxide - The C-O-O-C linkage. This is limited to peroxide in ring configuration only. The -C-O-O-H group is coded as "Misc." in 2-A. The peroxide descriptor does not include "ozonide" which is coded as O-hetero-fused-misc. E.g.: the peroxide group of



(ergosterol peroxide)

Figure 20

is coded 5, 8 O-hetero, peroxide, fused.

Pyranyl - includes saturated and unsaturated forms

Spiro, Fused and Independent - these terms refer to the manner in which the heterocycle ring is attached to a carbon atom or atoms of the steroid nucleus.

Miscellaneous - O-hetero substituents not specifically provided for above but specifically excluding epoxides, ketals, and acid and anhydride adducts.

13. Hal -

Halogen is coded wherever it appears except in acid halides (see COOR). It is further defined in 2-B (F, Cl, Br, I).

14. -S(Se, Te)-R

Any substituent joined to the steroid nucleus by attachment through a sulfur, selenium, or tellurium atom. S - containing heterocyclics are excluded.

It is further defined in 2-B as follows:

-Se-R The substituent is joined to the steroid nucleus through a selenium atom.

-Te-R The substituent is joined to the steroid nucleus through a tellurium atom.

=S(Se,Te) =S, =Se, or =Te is the substituent, and is attached by a double bond.

R=H Used for -S-H group.

R=other This includes all -S-R groups where R is not H, including $-SO_3H$, $-SO_2H$, etc.

In using the exo 2-A term =S(Se,Te) is specifically excluded where the =S(Se,Te) is part of an acyl group. Also specifically excluded are NCS and SCN which are provided for elsewhere.

15. S-het Ring

This designation refers to any sulfur-containing heterocyclic group attached to any of the 1 to 23+positions of the steroid nucleus through any of the atoms of the heterocycle.

The same general rules for coding rings are followed that were recited under O-hetero.

This category is further defined in 2B as follows:

Thiophene - the S-hetero in thiophene configuration, saturated or unsaturated.

Thiazole - the S-hetero in thiazole configuration, saturated or unsaturated. (Also coded as N-hetero.)

Thioketal - includes both cyclic monothioketals and cyclic dithioketals. The 2-A S-hetero descriptor is not used for "thioketal." The appropriate position is coded under "ketal."

Spiro, Fused, and Independent have been defined above. (See O-hetero.)

	GENERAL										HETERO ELEMENT OR HETERO RING													
	Non Substituent								Subst.		Oxygen													
	1 =		2 -H		3 allo, α		4 β		5 Misc.		6 -OH		7 =O		8 -O-Acyl		9 -O-R		10 Epoxy		11 Ketal		12 O-het ring	
12	Ex	21		21		21		21		21	Ex	21	Ex	21	Ex	21	Ex	21	Ex	21	Ex	21	Ex	21
11	23+	20	23+	20	23+	20	23+	20	23+	20	23+	20	23+	20	23+	20	23+	20	23+	20	23+	20	23+	20
0	22		22	10	22	10	22	10	22	10	22	10	22	10	22	10	22	10	22	10	22	10	22	10
1	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11
2	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12
3	3	13	3	13	3	13	3	13	3	13	3	13	3	13	3	13	3	13	3	13	3	13	3	13
4	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14
5	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15
6	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16
7	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17
8	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18
9	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

1 =

0	1(2)
1	1(10)
2	5(6)
3	5(10)
4	8(9)
5	8(14)
6	9(10)
7	9(11)
8	13(14)
9	13(17)
11	20(21)
51 12	20(22)
52 0	Aromatic A
1	Sat.ring A
2	Sat. ring B
3	Sat. ring C
4	Sat. ring D

2 Unsubstituted

5	Ring A
6	Ring B
7	Ring C
8	Ring D

8 -O-Acyl

0	Carboxylic
1	Poly
2	Unsat.
3	Subst.
4	Aromatic
5	Aliphatic
6	St. chain
7	Cyclo-al
8	Branched
9	Heterocyclic
11	Inorganic acyl
53 12	Phosphorus A
54 0	SO_4 (all S)
1	Osmium
2	Boron
3	Aliphatic
4	Aromatic
5	N containing
6	S containing
7	O containing
8	Halo containing
9	Misc.

9 -O-R

0	R=hydrocarbon
1	R=N group
2	R=S group
3	R=O group
4	R=Hal group
5	R=Other

12 O Hetero ring

6	Morpholine
7	Furan
8	Lactone
9	Spirostane
11	Sub in O spiro
55 12	Pseudosapo.
56 0	Acetonide
1	Peroxide
2	Pyranyl
3	Spiro
4	Fused
5	Independent
6	Misc.(N.E.K.)

13 Hal

7	F
8	Cl
9	Br
11	I

14 -S(SeTe)-R

0	-Se-R
1	-Te-R
2	=S(SeTe)
3	R=H
4	R=Other

15 S Het Ring

5	Thiophene
6	Thiazole
7	Thioketal
8	Spiro
9	Fused
11	Independent
12	Misc.
57	
58	

16 N-R-R

0	Primary amine
1	Secondary amine
2	Tertiary amine
3	Quaternary amine
4	NO_2, NO & Misc.
5	R=acyl
6	Imino

17

0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
11	
12	
59	
60	

18

0	
1	
2	
3	
21	
4	
5	
6	
7	
8	
9	

ED TO NUCLEUS									C OR C-RING ATTACHED TO NUCLEUS													
Non Oxygen									Hydrocarbon Substituent								Hetero Element in Chain					
	14 -S(SeTe) -R		15 -S het ring		16 -N-R-R		17 -N het ring		18 Keto reagt		19 CH_3		20 -C≡C-		21 HC Chain		22 HC Ring		23 -CN		24 -COOR		25 -C-R (-C subs)	
211	Ex	21	Ex	21	Ex	21	Ex	21	Ex	21		21	Ex	21	Ex	21		21	Ex	21	Ex	21		21
200	23+	20	23+	20	23+	20	23+	20	23+	20			23+	20	23+	20	23+	20	23+	20	23+	20	23+	20
100	22	10	22	10	22	10	22	10	22	10			22	10	22	10	22	10	22	10	22	10	22	10
111	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11
122	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12	2	12
133	3	13	3	13	3	13	3	13	3	13	3		3	13	3	13	3	13	3	13	3	13	3	13
144	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14	4	14
155	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15	5	15
166	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16	6	16
177	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17	7	17
188	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18	8	18
199	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19	9	19
266	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50

Riing

nee
nee
nee
iddine
lee
Lee
lennt
ntt
e
baizone
baigent
(1-7)
(8+)
ittic

22 HC Ring

0	3 M
1	4 M
2	5 M
3	6 M
4	7 + M
5	Sat
6	Unsat (N.Arom)
7	Aromatic
8	Spiro
9	Fused
11	Independent

61
62

24 COOR

0	R=H
1	R=salts (Met,Am)
2	R=alkyl
3	$C(=O)-X$ (X=hetero)
4	$C(=X)-O$ (X=hetero)

25 -C-R

5	R=N group
6	R=S group
7	R=O group
8	R=Halogen group
9	R=Other
11	Oxalyl

M Carbons at 17

0	0,1. (Androstane)
1	0 C
2	1 C
3	2 C (Pregnane)
4	21 Unsubst.
5	3 C
6	4 C
7	5 C
8	6+ C generic
9	6 C
11	7 C
63 12	8 C
64 0	9 C
1	10+ C
2	no 3-OR
3	Bile acids (3-5)
4	Bile (N.A.) (3-5)
5	Vitamin D
6	Isopregnane ≠ C-C

M Bis Subst. (Same)

7	At C(17)
8	At same C(not 17)

M General

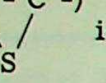
0	Addition
1	Maleic adduct
2	CNO, NCO, NCS, SCN
3	21 Diazo
4	Radioactive
5	i compounds
6	Microbiological
7	Oxidation
8	Reduction
9	Unsaturation
65 11	Ext. of natural materials

70 P.O. Classification
Class 260, subclass:

1	239.5	7	397.25
2	239.55	8	397.3
3	239.57	9	397.4
4	397	0	397.45
5	397.1	11	397.47
6	397.2	12	397.5

SERIAL NO.	Cl	DOCUMENT IDENTIFICATION

66 67 68 69 70 71 72 73 74 75 76 77 78 79 80

Miscellaneous - S-hetero substituents not specifically provided for above. The thiirane ring  is coded here.

16. N-R-R

This defines a nitrogen containing group connected through its nitrogen atom to the steroid nucleus. The term is not applicable when the nitrogen atom is part of a heterocyclic ring.

The 2-B descriptors for N-R-R are:

Primary, Secondary, Tertiary and *Quaternary Amines* are self-explanatory. E.g.:

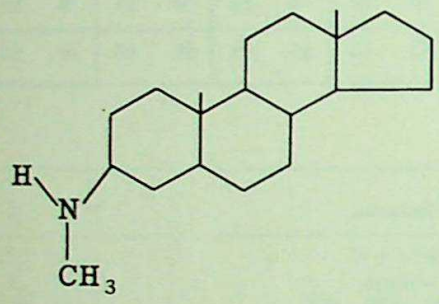


Figure 21

is coded 3- NRR in 2-A and "secondary amine" in 2-B.

NO_2, NO and Misc. - This term includes NO_2, nitroso and all NRR groups not covered by the other definitions in 2-B (e.g., nitrone, nitrate).

R = Acyl - converts the amine group to amides.

Imino - defines an imino group.

NOTE: =N - Ketone reagents are specifically excluded from N-R-R.

17. N-hetero Ring

This designation refers to a nitrogen-containing heterocyclic group attached through any of the atoms of the heterocyclic group.

The heterocycle can be fused, independent, or spiro. The rules set forth under O-hetero are followed.

The 2-B terms further define as follows:

Morpholine - (also coded as O-hetero)

Piperidine -

Pyridine - (including the dihydro and tetrahydro forms)

Pyrimidine - (including all saturated forms)

Azole - A 5-membered saturated or unsaturated ring containing at least one nitrogen. This is a generic term and all azoles are coded here.

Pyrrolidine - (Pyrrole) (including all saturated and unsaturated forms) - NOTE: Azole is also coded whenever this descriptor is used.

Thiazole - (including the saturated and unsaturated forms) - NOTE: Azole is also coded whenever this descriptor is used.

Piperazine -

Spiro, Fused and Independent have been defined above (see O-hetero)

Miscellaneous - N-hetero substituents not specifically provided for by the terms above.

18. Keto Reagent

This designation refers to a reaction product of a ketone or aldehyde group attached to any of the positions 1 to 23+ with well-known keto reagents.

It is further defined in 2-B as follows:

Hydrazone, Oxime and Semi-Carbazone are self-explanatory. Thiosemicarbazone is coded as semicarbazone and exo "S." When uncommon exo groups are present they are recorded (e.g., in dinitrophenyl hydrazone the nitro groups would not be coded; but in di-cyano phenylhydrazone the cyano groups would be coded). The above should be considered when attempting to find substituted keto reagents.

Girard Reagent - includes all acyl hydrazones substituted on the acyl moiety by a quaternary ammonium radical. Girard reagents are not coded as hydrazones and are not split into "exo" terms, except for halogen.

NOTE: Substituents coded here are not coded under NRR.

19. CH_3

This represents a methyl group. This code is not used at positions 10, 13, 20, 22, or 23+ because methyl groups in these positions are considered integral parts of the steroid nucleus. When a methyl group is coded, the 2A term "hydrocarbon chain" and the 2B term "lower alkyl (1-7)" are also coded. The "exo" - CH_3 descriptor is not used.

20. $-C\equiv C-$

This represents the ethinyl linkage. When this code is used "hydrocarbon chain" in 2-A and "alkinyl" in 2-B are also coded.

21. HC (Hydrocarbon) Chain

This term represents a hydrocarbon chain. Methyl groups at the 10 and 13 positions are not coded (see the -CH_3 descriptor). Also, in pregnanes, the chain of 2 carbons attached to the 17 position is not coded as a hydrocarbon

unless it is unsubstituted (i.e., $-CH_2-CH_3$). In cholesterols and other sterols, the hydrocarbon chain in the 17, 20 and 22 positions is not coded as hydrocarbon.

Hydrocarbon chain includes aralkyl or any other cyclic hydrocarbon ring attached to the steroid nucleus through an aliphatic carbon chain but the carbons of the ring are not counted as chain members.

This descriptor is not used at positions 20, 22, or 23+ (except in the case of 20-, 22-, or 23+ - ethinyl) since carbons attached to these positions are integral parts of the steroid nucleus.

"Exo"-HC Chain is not coded except when used in conjunction with the "exo"-C≡C- descriptor.

The 2-B definitions are:

L. Alkyl (1-7) a chain of 1 to 7 carbons

Hi. Alkyl (8+) a chain of 8 or more carbons

$=CH_2$ this includes both H and unsubstituted hydrocarbon substituents attached to $=C<$.

Alkenyl - if the HC chain contains a double bond other than at the point of attachment to the steroid nucleus this code is used.

With Aromatic - an aliphatic side chain substituted by an aromatic hydrocarbon ring.

Alkinyl - if the HC chain contains a triple bond this descriptor is used.

22. Hydrocarbon Ring

This term includes any hydrocarbon ring including those having non-hydrocarbon substituents, attached to the nucleus in positions 1-23+. In the case where the ring is fused to the nucleus the positions on the ring to which it is attached are recorded. The "exo"-HC ring descriptor is not used.

The 2-B definitions are:

3 M, 4 M, 5 M, 6 M, and 7 M all denote the number of carbon atoms in the ring (M-members).

Sat - used if the ring is saturated.

Unsat (N. Arom.) - used if the ring is unsaturated but not for aromatic rings.

Aromatic -

Spiro, Fused, and *Independent* are described under O-hetero.

23. CN

This represents the nitrile (cyano) group.

24. COOR

This term represents carboxylic acid radicals and their salts, esters, and amides. It also includes thioesters, thioamides, and halides. COOR is considered as a unit and the -OH, =O, $-NH_2$, etc. portions are not coded elsewhere. The 2-B terms are:

R=H - the carboxyl group

R=salts - metal and amine salts of the carboxyl group

R=alkyl - esters of the carboxyl group with any alcohol, including aromatic alcohols.

C - X, X=hetero - X includes any hetero atom, the most common of which are nitrogen (amides) or the halides.

C - O, X=hetero - most commonly X will be sulfur. (NOTE: in the case of dithio acids both of the latter codes are recorded).

25. -C-R (C-subs)

This symbol represents a non-hydrocarbon substituent linked through a carbon atom to the steroid nucleus and not specifically provided for by any other 2-A term.

In cases of doubt the C-sub is superior to "Miscellaneous" and if a group can be coded in C-sub, it is coded there rather than in "Miscellaneous."

Substituted cycloalkyls attached to the nucleus are coded HC-ring and not as C-sub. The substituting groups are coded as "exo."

The substituted carbon group is further defined in 2-B

R=N containing group
R=S " "
R=O " "
R=halogen " "
R=other

R=oxalyl. This defines the $-\overset{O}{\overset{\|}{C}}-\overset{O}{\overset{\|}{C}}-$ group.

(21-oxalyl is also coded 4-carbons at 17; the sodium enolate is coded 21- double bond and 21-C-sub, R=O containing group.)

NOTE: The exo C-sub descriptor is not used. In these cases the radical that is substituted on the carbon chain is coded as exo. Also C-sub is not employed at the 10, 13, 17, 20, 22 and 23+ positions. For example:

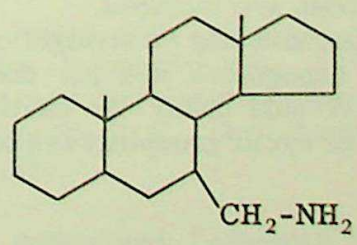


Figure 22

Code 7-C-sub in 2-A and N-cont. group in 2-B (as well as "exo" N-R-R and primary amine)

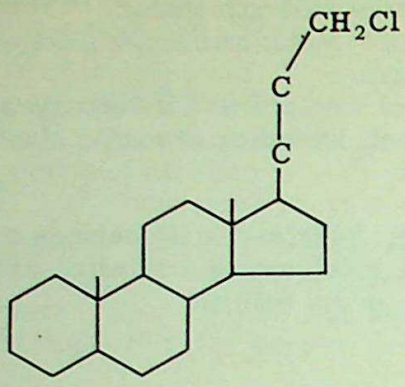


Figure 23

Code 21-C-sub in 2-A and "halogen group" in 2-B. (Also code "exo" halo, chlorine, and 3 carbons at 17.)

The foregoing 25 categories are all 2-A terms, some of which were further defined in 2-B.

CODING OF 2-B TERMS

The following are found only in the 2-B section and are not specific to any of the 2-A terms.

Unsubstituted

Ring A - no substituents at any of the positions 1, 2, 3, 4, 5, and 10 in any one of the compounds of the document. (10-methyl is not a substituent.)

Ring B - positions 6 and 7

Ring C - positions 8, 9, 11, 12, 13, and 14 (13 methyl is not a substituent).

Ring D - positions 15, 16, and 17 (the 17-side chain in pregnanes, sterols, etc., is considered a substituent).

Carbons at 17

Only the largest number of carbons in a single group at 17 is counted. Carboxy and cyano carbons are included.

The side chain may be straight or branched. Steroidal sapogenins are not considered as having a 17-side chain. No carbons that are members of cyclic groupings are counted here. E.g.:

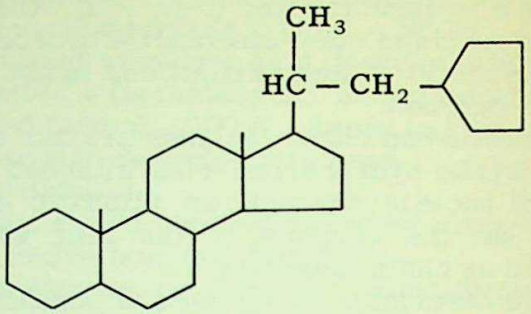


Figure 24

is coded as 3 carbons at 17.

NOTE: When 63-0 is coded, one of 63-1 and 63-2 must also be coded. Also, when 63-4 is coded, 63-3 must be coded. When any of 63-9, 11, or 12 or 64-0 or 1 are coded there must be an accompanying 63-8 code.

0, 1	*(Androstane)*	*is self-explanatory*	
0	*C*	"	"
1	*C*	"	"
2	*C (Pregnane)*	"	"

Note: 21 Norprogesterone is coded as a pregnane (also as 20-keto).

	21 Unsubst.	*is self-explanatory*	
3	*C*	"	"
4	*C*	"	"
5	*C*	"	"
6 +	*C generic*	"	"
6	*C*	"	"
7	*C*	"	"
8	*C*	"	"
9	*C*	"	"
10 +	*C*	"	"

6+ no 3-OR on nucleus - this term represents sterols having no substituent at the 3-position or some substituent other than oxygen at said position.

Bile acids (3-5) (includes bisnorcholanic, norcholanic and cholanic acid side chains)

Bile (N.A.) (3-5) All compounds which have a chain of 3-5 carbon atoms in bile acid configuration in the 17 position but definitely excludes the acid. Bile (Non Acid) contains one of the following groups but excludes everything which has been defined under COOR previously:

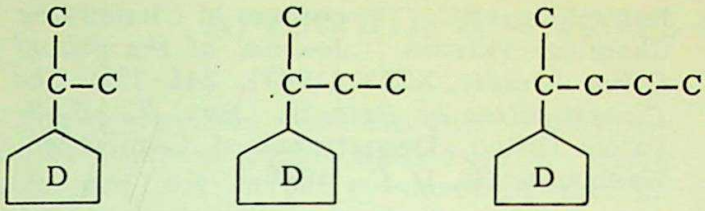
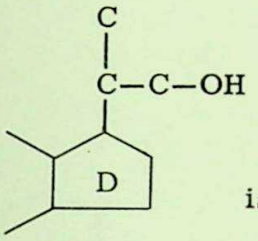


Figure 25

e.g.,



is coded Bile N.A.

Figure 26

Vitamin D - all members of the Vitamin D family are included even though they are not steroids.

Isopregnane α C-C is coded when there is a saturated, substituted 2 carbon chain in alpha configuration at the 17 position.

Bis Subst. (Same)

This is used when a steroidal carbon is attached to two identical groups. It is not used to record bismethyl substitutions in the side chain as in the case of the 26 and 27 methyl groups of cholesterol nor is it used for hydrogen atoms or for symmetrical spiro rings (e.g., ketals). It is further divided into two specific descriptors

At C(17)

At same C (not 17)

Example: lanosterol is coded bis "at same C (not 17)" for the 4, 4 bismethyl groups.

M General (Misc.)

Addition - This term designates addition compounds such as bisulfate addition produces, hydrates, Grignard addition compounds, etc. (Also code under 2-A Miscellaneous when possible.)

Maleic adducts - Steroid reaction products of the type of maleic acid anhydride or ester adducts. Note: Maleic adducts take precedence over all other codes and are only coded as such except for nuclear substituents. No other codes are used for the adduct.

CNO, NCO, NCS, SCN - Self-explanatory (also code in 2-A under "Miscellaneous").

21 Diazo - Self-explanatory (not coded 21-N-R-R or 21-Misc.).

Radioactive - This includes all steroids that are radioactive or have radioactive isotopes as part of their structures.

i compounds - Self-explanatory, not considered as a fused compound.

Microbiological - This describes a chemical process carried out by microbiological methods.

Oxidation, Reduction and Unsaturation are specific under "Microbiological" for particular processes. Both generic and specific terms are coded.

Ext. of natural materials is used whenever the document recites a natural material extraction process.

ALTERNATIVE CODING PROCEDURE— NON-COMPOSITE METHOD

A modification of the coding system described above can be employed when it is desired to provide machine selection and discrimination on an individual compound basis. This modification is in contrast with the method of composite coding.

In individual compound coding, the 2A substituents and 2B terms disclosed for the particular compound are coded in the usual manner. The absence of 2A terms in the remaining positional locations on the steroid nucleus is indicated by employing the "H" descriptor for each of such locations. The only exception is in the 17 position, in which the punching of an "H" signifies the presence of only one substituent instead of two. When the keto group is in positional location 17, no "H" is punched. This device is not used for positional locations 20, 21 and 22.

To find those compounds which do not have double bonds in certain positions, the absence of a double bond is asked for by an appropriate wiring modification.

REFERENCES

1. Frome, Julius and Leibowitz, Jacob. *A Manual for Coding Steroids*. Patent Office Research and Development Report No. 11. Washington 25, D. C., Department of Commerce, 1958.

2. Frome, Julius and Leibowitz, Jacob. *A Punched Card System for Searching Steroid Compounds*. Patent Office Research and Development Report No. 7. Washington 25, D. C., Department of Commerce, 1956.

3. *Classification Bulletin, Class 260, Chemistry, Carbon Compounds*. No. 200, Rev. 1. U. S. Patent Office, Department of Commerce, Washington 25, D. C., 1956.

4. *Ibid.*, pp. 260-37, 260-38, 260-68 through 260-70.

5. Rosa, Manuel C. "Problems of Classifying Chemical Patents." *Journal of the Patent Office Society*, XXIX (1947), 241-261; *The Classification of Patents*. Rev. 2. U. S. Patent Office, Department of Commerce, Washington 25, D. C., 1946.

APPENDIX A - NOTES ON THE USE OF "EXO"

The designation "Ex" represents substituents or pairs of double bonded carbon atoms which are not directly connected through carbon-to-carbon linkages to the 1-23+ carbon atoms of the steroid nucleus. For example, the carboxyl substituent in cholic acid

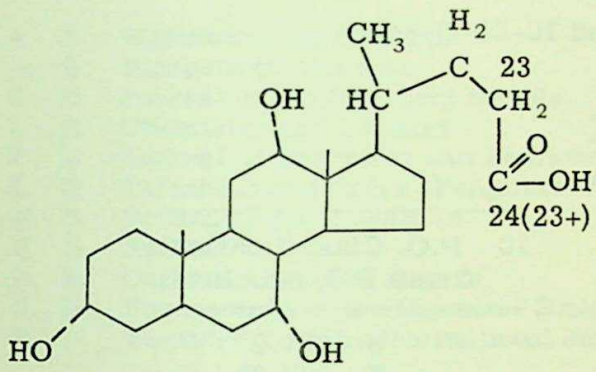


Figure 27

is directly connected, and is coded as 23+ COOR. The sulfo radical in taurocholic acid

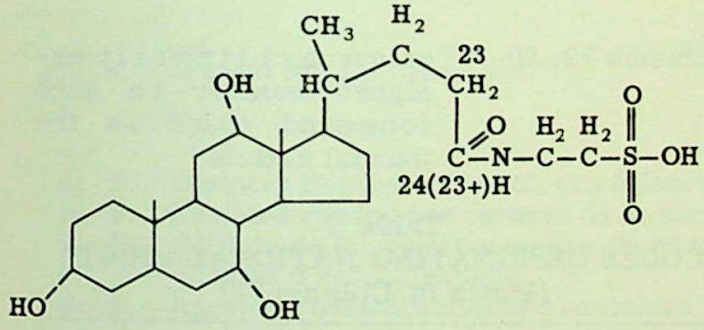


Figure 28

and the carboxyl radical in glycholic acid

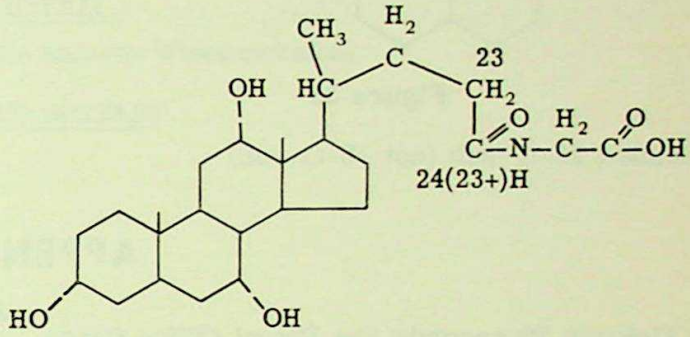
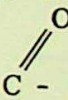


Figure 29

are not directly connected, and are coded as Ex-S-R and Ex-COOR respectively.

Note that in each of these compounds, the 2A term 23+ COOR and the 2B term



X=hetero are also coded to represent the carboxylic acid derivative substituents at the 23 positions.

The designation "Ex" is *not coded* in the following 2A columns:

H ; α ; β ; Misc. ;

CH_3 ; HC Ring ; HC Chain* ; -C- sub .

*<u>But Note</u>: alkinyl - C≡C - is coded as an "Ex" term, and when it is, HC Chain is also coded.

APPENDIX B - NOTE ON CODING AT 10, 13, 18 AND 19

When the methyl groups at positions 10 and 13 are substituted the substituents are coded from positions 18 and 19 unless these positions are part of a ring in which case they are coded from positions 10 and 13.

A few examples will serve to illustrate this:

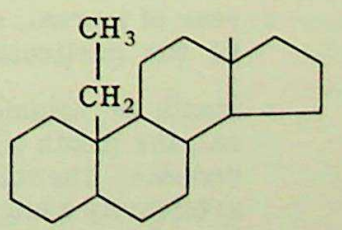


Figure 30

is coded 19-methyl (not 10-HC chain)

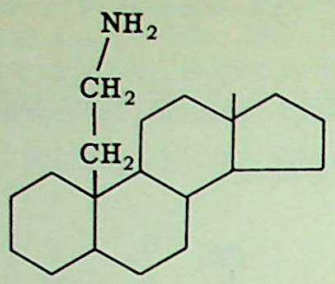


Figure 31

is coded 19-C-sub (not 10-C-sub)

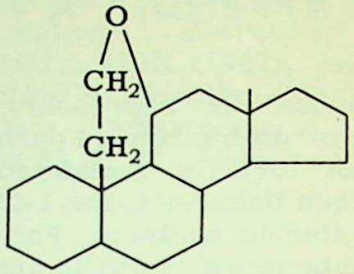


Figure 32

is coded 10-11-O-hetero

APPENDIX C

Column 70 records the Patent Office Classification of patents coded. (See reference 3.)

70	P.O. Classification Class 260, subclass:
1	239.5
2	239.55
3	239.57
4	397
5	397.1
6	397.2
7	397.25
8	397.3
9	397.4
10	397.45
11	397.47
12	397.5

APPENDIX D

Column 78 : code designating the country of patent origin; see Table I for particular codes.

Columns 71-77 : patent number.

For published literature, the remaining ten columns are allocated as follows:

Column 71 : code representing the journal.

Column 72 : code designating the country of journal origin; see Table II for the combined journal—country codes.

Column 73 : year of journal; see Table III for particular codes.

Column 74 : month or volume of journal, the month taking precedence. The volume is an arbitrarily assigned number—1 for 1958, 2 for 1959, etc.

Columns 75-78 : page numbers of journal article.

Columns 79, 80 : unique arbitrarily assigned number to each compound coded in the journal article.

Table I
CODES DESIGNATING NATIONAL ORIGIN
(Punch in Column 72)

Code	Nation
A	Austria
B	Great Britain
C	Canada
D	Denmark
E	India
F	France
G	Germany
H	Switzerland
I	Italy
J	Japan
K	Australia
L	China
N	Netherlands
Q	Spain
R	Russia
U	United States
Z	Czechoslovakia

Table II

REVISED LISTING OF CODES IDENTIFYING JOURNAL AND NATIONAL ORIGIN

(NOTE: This revised list includes all corrections and additions to date. Reference to previous lists should be ignored.)

A - AUSTRIA

C	A	Monatshefte fur Chemie und Verwandte Teile anderer Wissenschaften

B - GREAT BRITAIN

A	B	Manufacturing Chemist
B	B	Biochemical Journal
C	B	Journal of the Chemical Society
I	B	Chemistry and Industry
J	B	*Journal of Pharmacy and Pharmacology
L	B	Tetrahedron Letters (Pergamon Press Ltd.)
M	B	British Medical Bulletin
N	B	Nature
P	B	Current Chemical Papers
S	B	Proceedings of the Chemical Society of London
T	B	Tetrahedron (The International Journal of Organic Chemistry) Pergamon Press

C - CANADA

B	C	Canadian Journal of Biochemistry and Physiology
C	C	Canadian Journal of Chemistry

D - DENMARK

A	D	Acta Endocrinologica
C	D	Acta Chemica Scandinavica

E - INDIA

C	E	Journal Indian Chem. Soc.

F - FRANCE

A	F	Comptes Rendus de l'acad. des Sciences
B	F	Comptes rendus des seances de la societe de biologie et de ses filiales
C	F	Bulletin de la societe chimique de France
E	F	Annales d'endocrinologie
P	F	Annales Pharmaceutiques Francaises

G - GERMANY

B	G	Chemische Berichte
C	G	Angewandte Chemie
F	G	Zeitschrift fur Naturforschung
L	G	Annales der Chemie, Justus Liebigs
N	G	Naturwissenschaften
P	G	Zeitschrift fur physiologische Chemie (Hoppe-Seyler's)
Z	G	*Zeitschrift fur chemie

H - SWITZERLAND

A	H	*Chimia
C	H	Helvetica Chimica Acta
E	H	Experientia

Table II—REVISED LISTING OF CODES IDENTIFYING JOURNAL AND NATIONAL ORIGIN—Con.

I - ITALY

A	I	Annali Di Chimica
C	I	Gazetta Chimica Italiana
F	I	*Il Farmaco Edizione Scientifica

J - JAPAN

A	J	Proceedings Japan Academy
B	J	Journal of Biochemistry
C	J	Journal of the Chemical Society of Japan
D	J	Journal of the Agricultural Chemical Society of Japan
F	J	Agricultural and Biological Chemistry (formerly Bulletin of the Agricultural Chemical Society)
H	J	Hiroshima Journal of Medical Science
J	J	Journal of the Pharmaceutical Society of Japan
M	J	Yonago Acta Medica
P	J	Chemical and Pharmaceutical Bulletin (formerly Pharmaceutical Bulletin)
S	J	Bulletin Chemical Society

K - AUSTRALIA

C	K	Australian Journal of Chemistry

L - CHINA

C	L	Acta Chimica Sinica
S	L	Scientia Sinica-Academia Sinica

M - BELGIUM

C	M	*Bulletin des Societies Chimiques Belges

N - NETHERLANDS

C	N	Recueil des Travaux Chimiques des Pays-Bas

Q - SPAIN

P	Q	Anales de Fisica y Quimica

R - RUSSIA

D	R	Doklady Akademii Nauk Soyuza Sovetskikh Sotsialisticheskikh Respublik
I	R	Izvestiya Akademii Nauk Soyuza Sovetskikh Sotsialisticheskikh Respublik Otdelenie Khimicheskikh Nauk (Classe des sciences chimiques)
M	R	Voprosy Med. Khimii
O	R	Zhurnal Obshchei Khimii (Journal Gen. Chem. USSR)
P	R	*Meditsinskaya Promishlennost SSSR
U	R	Ukrainskii Khimicheskii Zhurnal

U - UNITED STATES

A	U	Archives of Biochemistry and Biophysics Academic Press
B	U	Journal of Biological Chemistry
C	U	Journal of the American Chemical Society
D	U	Endocrinology
E	U	Journal of Clinical Endocrinology and Metabolism
F	U	Journal of the American Pharmaceutical Association
G	U	*Chemical and Engineering News
J	U	Journal of Pharmaceutical Sciences
M	U	Proceedings of the Society for Experimental Biology and Medicine
N	U	Proceedings of the Federation of American Societies for Experimental Biology

Table II—REVISED LISTING OF CODES IDENTIFYING JOURNAL AND NATIONAL ORIGIN—Con.

U - UNITED STATES—Con.

O	U	Journal of Organic Chemistry
P	U	American Journal of Physiology
R	U	Biochemical and Biophysical Research Communications
S	U	Science

Z - CZECHOSLOVAKIA

C	Z	Collection of Czechoslovak Chemical Communication

*These are journals which we do not regularly receive. However, whenever articles in these journals are found or called to our attention, we code them.

Table III
CODES DESIGNATING YEAR OF JOURNAL
(Punch in Column 73)

Year	Code	Year	Code
1958	11	1964	4
1959	12	1965	5
1960	0	1966	6
1961	1	1967	7
1962	2	1968	8
1963	3	1969	9

34117--U.S.Dept.of Comm--DC--1962